Inequalities For Random Multilinear Operators

Michael Christ, UC Berkeley

BIRS, July 26, 2010

Three Themes

- Fourier transform and linear structure
- Multilinear operators
- Probability/random structures

Bilinear operators aka trilinear forms

$$\mathcal{T}_\omega(f,g,h) = \sum_{x,y=1}^{N} f(x)g(y)h(x+y)k_\omega(x-y)$$

where k_ω is a random probability measure.

One point of view re multilinear operators:
$\mathcal{T}_\omega(f,g,h) =$ inner product of $T_{\omega,h}f$ with test function g. Seek worst case (in h) bounds for **linear operator** $T_{\omega,h}$.

Worst case inequalities comparing $\mathcal{T}_\omega(f,g,h)$ to its expected value, for **large** N — **worst** h**, for typical** ω.

A venerable theme: Smallness of Fourier transforms in absence of linear structure

- If μ is supported on a curved submanifold, then $\widehat{\mu}(\xi) = O(|\xi|^{-\rho})$ as $|\xi| \to \infty$.
- Let $\mu =$ random probability measure on $\mathbb{Z}_N$, $m =$ uniform probability measure. Then $\max_{\xi \neq 0} |\widehat{\mu}(\xi) - \widehat{\mathbf{m}}(\xi)| = \mathbf{O}(\mathbf{N}^{-1/2} \log(\mathbf{N}))$ with high probability.
- Natural Cantor-Lebesgue-type probability measures on random fractal sets have Fourier transforms which **tend to zero** at a natural rate as $|\xi| \to \infty$. (e.g. Salem 1951)
- Let p= large prime and $\mu_p(x) = 1$ if x is a **quadratic residue** modulo p, and $\mu_p(x) = 0$ otherwise. Then $\sup_{\xi \neq 0} |\widehat{\mu_{\mathbf{p}}}(\xi)| \leq \mathbf{C}\mathbf{p}^{-1/2}$, whereas $\widehat{\mu}_p(0) \asymp 1$.

One More Illustration

Consider random matrix

$$\begin{pmatrix} r_{1,1}(\omega) & \cdots & r_{1,N}(\omega) \\ \vdots & \vdots & \vdots \\ r_{N,1}(\omega) & \cdots & r_{N,N}(\omega) \end{pmatrix}$$

with entries which are: $O(N^{-1})$, iid, with **mean zero**.

With high probability as $N \to \infty$, the ℓ^2 operator norm is $O(\mathbf{N^{-1/2}} \cdot N^{\varepsilon})$.

Quantum Interpretation(s)

- Smallness of $\widehat{\mu}$ can be reinterpreted operator-theoretically in terms of $T(f) = f * \mu$, by virtue of Plancherel's theorem; Small Fourier transform $\Leftrightarrow$ small operator norm.
- **Goal:** Smallness of $\mathbb{C}$-valued multilinear form

$$\mathcal{T}_\omega(f_1, \cdots, f_M) = \sum_{x,y=1}^{N} \mathbf{r}_\omega(\mathbf{x}, \mathbf{y}) \prod_{j=1}^{M} f_j(\mathbf{L_j}(\mathbf{x}, \mathbf{y}))$$

where $L_j : \mathbb{Z}^2 \to \mathbb{Z}$ are **linear**
and $r_\omega(x, y)$ are either jointly **independent**
or (Toeplitz case)
$r_\omega(x, y) \equiv r_\omega^\heartsuit(\mathbf{x} - \mathbf{y})$ with $\{r_\omega^\heartsuit(x)\}$ jointly independent.

- We are interested in $\ell^{\mathbf{p_1}} \otimes \ell^{\mathbf{p_2}} \otimes \cdots \otimes \ell^{\mathbf{p_M}}$ bounds with $\sum_j p_j^{-1} = 1$, that is,

$$\Big| \sum_{x,y=1}^{N} r_\omega(x,y) \prod_{j=1}^{M} f_j(L_j(x,y)) \Big| \lesssim \mathbf{N}^{-\rho} \prod_j \|f_j\|_{p_j}$$

  where r_ω has **mean zero** and $\mathbb{E}|r_\omega| \asymp N^{-1}$.
- These "averaging" type bounds scale naturally for **Tauberian-style** **ergodic-theoretic** interpretations.
- The cancellation condition $\mathbb{E}_\omega(r_\omega(x,y)) = 0$ is essential for smallness of the operator norm.
- Mean zero arises naturally by comparing more general objects to their mean/expected values.

A multilinear inequality in terms of $\ell^2 \otimes \ell^2 \otimes \ell^\infty \cdots \otimes \ell^\infty$ is equivalent to a **worst case** estimate for a **linear** operator:

- Modify random matrix

$$\big(r_\omega(x,y)\big)_{x,y=1}^{N}$$

by multiplying entries by arbitrary

$$\prod_{k=3}^{M} f_k(L_k(x,y)) \text{ with } \|f_k\|_{\ell^\infty} \le 1.$$

- We want to bound the largest possible norm.
- If f_k were allowed to depend freely on both variables (x,y), then cancellation could be completely destroyed and the best estimate would be $O(1)$.

A Cautionary Example

Let $G_p = \mathbb{Z}_p^d \times \mathbb{Z}_p$; $|x'|^2 = |(x_1, \cdots, x_d)|^2 = \sum_{j=1}^d x_j^2$.

$\mu_{\mathbf{p}}(x', x_{d+1}) = \begin{cases} p^{-d} & \text{if } \mathbf{x_{d+1} = |x'|^2} \\ 0 & \text{otherwise} \end{cases}$,

$\nu_{\mathbf{p}} = \mu_p - p^{-d-1}$.

Then

$$\Big|\sum_{x,y} f(x)g(y)\nu_p(x-y)\Big| \lesssim p^{-d/2}\|f\|_2\|g\|_2 \;\forall f, g,$$

but there exist f, g, h such that

$$\Big|\sum_{x,y} f(x)g(y)\mathbf{h(x+y)}\nu_p(x-y)\Big| = \|f\|_2\|g\|_2\|h\|_\infty;$$

there is no cancellation at all.

The counterexample:

$$h(x) = e^{2\pi i|\mathbf{x}'|^2/p}$$
$$f(x) = e^{2\pi i[\mathbf{x_{d+1}} - \mathbf{2}|\mathbf{x}'|^2]/p}$$
$$g(x) = e^{2\pi i[-\mathbf{x_{d+1}} - \mathbf{2}|\mathbf{x}'|^2]/p}.$$

satisfy

$$f(x)g(y)h(x+y) \equiv 1 \text{ when } x_{d+1} - y_{d+1} = |x' - y'|^2$$

but not at typical points $(x, y) \in G_p^2$.

The lesson: An obstruction to the trilinear inequality is **quadratic structure** (of μ_p).

This issue is related to the distinction between uniformity and **Gowers uniformity**, which is at the heart of certain advances in additive combinatorics related to Szemerédi's theorem, but is not exactly the same issue.

First Theorem

Consider linear operator, with $\|\mathbf{g_j}\|_\infty \leq \mathbf{1}$:

$$T_{\omega,\{g_j\}}(f)(x) = \sum_{y=1}^{N} r_\omega(\mathbf{x}-\mathbf{y}) f(x) \prod_{\mathbf{j=1}}^{\mathbf{M}} \mathbf{g_j}(\mathbf{L_j}(\mathbf{x},\mathbf{y})).$$

Let: $\Omega =$ probability space with $\{s_\omega(x) : x \in \mathbb{Z}\}$ iid $\{0,1\}$–valued
$s_\omega(x) = 1$ **with probability** p
$r_\omega(x) = (Np)^{-1} s_\omega(x) - N^{-1}$ for integers $x \in [-N, N]$
Thus $\mathbb{E}_\omega r_\omega(x) \equiv 0$ for $x \in [-N, N]$
while $\mathbb{E}_\omega |r_\omega(x)| \asymp N^{-1}$.

Theorem

Suppose that $M \geq 1$ and $0 \leq \gamma < 2^{-M}$. There exists $\varepsilon > 0$ such that for all $N \geq 1$ and $p = N^{-\gamma}$,

$$\mathbb{E}_\omega \sup_{\{\mathbf{g_j}\}} \| T_{\omega,\{g_j\}} \|_{op} \leq C N^{-\varepsilon}.$$

Same Theorem — A Defect?

Theorem

Suppose that $M \geq 1$ *and* $0 \leq \gamma < 2^{-M}$. *Let* $p = N^{-\gamma}$. *Then*

$$\mathbb{E}_\omega \sup_{\{g_j\}} \| T_{\omega,\{g_j\}} \|_{op} \leq CN^{-\varepsilon}.$$

The theorem applies only when the matrix $\big(s_\omega(x-y)\big)$ is not too sparse, in terms of N; e.g. in the trilinear case, our proof requires that the density of points "selected" be $\gg N^{-1/2}$.

I simply **do not know** whether $\gamma < 2^{-M}$ is necessary. Method of proof does break down irretrievably past this threshold.
Could restriction be an artifact of the proof? Today's results should be regarded as preliminary.

An Easier Result

- Order of quantifiers matters.

$$\sup_{g_1,\cdots,g_M} \mathbb{E}_\omega \sup_f \| T_\omega(f, g_1, \cdots, g_M)\|_2$$

is a related, but possibly smaller, quantity.

- Easier result:

$$L_{\omega,h}(f)(x) = \sum_y r_\omega(x-y)\ \mathbf{h(x,y)}\ f(y)$$

satisfies

$$\mathbb{E}_\omega \|L_{\omega,h}\|_{\text{op}} \le C_\varepsilon N^\varepsilon (Np)^{-1/2} \|h\|_{\ell^\infty}$$

for all $\varepsilon > 0$ provided $p \ge N^{-\gamma}$ and $\gamma < 1$.

- Proof: Expand a high power of $L^*_{\omega,h} L_{\omega,h}$ and take expectation of its **trace**.

Carleson-style maximal analogue

$$T^*_{\omega,\{g_j\}}(f)(x) = \sup_{\xi} \Big| \sum_{y=1}^{N} \mathbf{e^{i\xi y}} r_\omega(x-y) f(x) \prod_{j=1}^{M} g_j(L_j(x,y)) \Big|.$$

Theorem

Suppose that $M \geq 1$, *and* $p = N^{-\gamma}$ *where* $0 \leq \gamma < 2^{-M-1}$. *Then*

$$\mathbb{E}_\omega \sup_{\{g_j\}} \| T^*_{\omega,\{g_j\}} \|_{op} \leq C N^{-\varepsilon}.$$

Application to Ergodic Theory

Let $T =$ invertible measure-preserving transformation on probability space.
Let $s_n(\omega) = 1$ with probability $n^{-\gamma}$, and $= 0$ otherwise.
Random sparse subsequences of $\mathbb{N}$: $(n_k(\omega))_{k\in\mathbb{N}}$ consists of all $n \in \mathbb{N}$ for which $s_n(\omega) = 1$, listed in increasing order.

Theorem

If $0 \le \gamma < 2^{-M+1}$ then for almost every $\omega \in \Omega$, for all $f_1, \cdots, f_M \in L^\infty(X)$,

$$\lim_{N\to\infty} N^{-1} \sum_{k=1}^{N} f_1(T^{n_k}(x)) f_2(T^{2n_k}(x)) \cdots f_M(T^{Mn_k}(x))$$

exists in $L^1(X, d\mu(x))$.

Another Application to Ergodic Theory

- For the full sequence of iterates in the theorem on the preceding slide, see Tao and Host-Kra, also an alternative approach of Austin. Theirs is the deep result; the refinement to subsequences is a comparatively simple add-on.
- The Carleson-style maximal analogue has a corresponding application to an extension of the *Return Times* theorem (Bourgain; Demeter-Lacey-Tao-Thiele), replacing averages over a full sequence of iterates by averages over a sparse random subsequence of iterates.

Non-Toeplitz-style Variant

- Next: Analogous results for random matrices $\big(r_\omega(x,y)\big)_{x,y}$, with all entries jointly independent.
- Consider jointly independent random selector variables $s_\omega(x,y)$ for $(x,y) \in [-N,\cdots,N]^2$, satisfying $s_\omega(x,y) = 1$ with probability p, and $= 0$ otherwise.
- Then $\mathbb{E}(\sum_x s_\omega(x,y)) \asymp Np$ and $\mathbb{E}(\sum_y s_\omega(x,y)) \asymp Np$.
- Define $r_\omega(x,y) = (Np)^{-1}\big(s_\omega(x,y) - p\big)$ so that $\mathbb{E}_\omega r_\omega(x,y) = 0$.

Non-Toeplitz-style Variant

Consider

$$T_{\omega,\{g_j\}}(f)(x) = \sum_y r_\omega(x,y) f(y) \prod_{j=1}^{M} g_j(L_j(x,y)).$$

As always, $\|g_j\|_\infty \le 1$.

Theorem

Let $M \ge 1$ and $0 \le \gamma < 1$. For $N \ge 1$ set $p = N^{-\gamma}$. For any $\{L_j : 0 \le j \le M\}$ and any $\varepsilon > 0$,

$$\mathbb{E}_\omega \sup_{\{g_j\}} \|T_\omega\|_{op} \le C_{M,\varepsilon} N^\varepsilon N^{-(1-\gamma)/2}.$$

Method of Proof

For $T(f)(x) = \sum_y r_\omega(\mathbf{x} - \mathbf{y}) f(y) g(x+y)$:

- $$\|Tf\|_{\ell^2}^2 = \sum_{z \in \mathbb{Z}} \Big(\sum_{x,y} F_z(x) G_z(x+y) \rho_{\omega,z}(x-y) \Big)$$

  where $F_z(x) = f(x) f(x+z)$,
  $G_z(x) = g(x) g(x+z)$,
  $\rho_{\omega,z}(x) = r_\omega(x) r_\omega(x+z)$.
- Fix arbitrary z. After linear change of variables, inner sum represents a **linear convolution operator** $\ell^2 \to \ell^2$.
- Need bound for $\widehat{\rho_{\omega,z}}$.

Issues

- Must sacrifice a factor of $N^{1/2}$ to control $\|G_z\|_{\ell^2}$ in terms of $\|g\|_{\ell^\infty}$.
- $\rho_{\omega,z}$ is a product of two singular measures, hence is even more singular.
- Need bounds for $\widehat{\rho_{\omega,z}}(\xi) = \sum_x r_\omega(x) r_\omega(x+z) e^{-ix\xi}$.
- **Independence of summands no longer holds.**
- $\widehat{\rho_{\omega,z}}$ is very badly behaved for $z = 0$. But a bounded number of exceptional z can be handled by a different (trivial) bound.
- **If our original measure is too sparse**, then the support of $\rho_{\omega,z}$ may consist of one or zero points for most z. Then there will be no possible cancellation in the calculation of $\widehat{\rho_{\omega,z}}$. The argument then breaks down utterly.

Proofs, continued

- Higher degrees M of multilinearity are treated by induction.
- TT^* is applied as above, but repeatedly; each application reduces M by 1.
- **Different base case** for different M. Linear convolution operator, with $r_\omega(x)$ replaced by $\prod_{j=1}^{M} r_\omega(x+z_j)$ for arbitrary $(z_1, \cdots, z_M)$.
- Each iteration leads to a small number of exceptional parameters z, which must be handled differently.
- For large M, the product of M translates of r_ω is very singular.

Proof for non-Toeplitz case $r_\omega(x,y)$

- Suppose f, g, h are characteristic functions of sets F, G, H. **Fix F, G, H.**
- Our trilinear form is $\sum_{(x,y)\in\mathcal{E}} r_\omega(x,y)$ where $\mathcal{E} = \{(x,y) : x \in F,\ y \in G,\ \text{and } x+y \in H\}$.
- An auxiliary argument reduces matters to the case where $|\mathcal{E}| \gtrsim N^{2-\eta}$ for a natural (and small) value of η.
- This is a sum of $|\mathcal{E}| \gg 1$ independent random variables, so is **within a bounded number of standard deviations of its mean $(= 0)$ with high probability**. Its standard deviation is proportional to

$$N^{-1}p^{-1/2}|\mathcal{E}|^{1/2} \asymp p^{-1/2} \ll N^{1/2},$$

while the bound we seek is

$$N^{1-\varepsilon}.$$

Entropy

- Thus a standard Gaussian distribution would give the probability of a bad event, for fixed F, G, H, to be

$$\lesssim e^{-cN^{1+\delta}}$$

for a certain $\delta(\gamma) > 0$, if $p \asymp N^{-\gamma}$ with $\gamma < 1$.

- Chernoff's inequality (a generalization of Khinchine's inequality) gives hybrid exponential/exponential squared large deviations bound which suffices for this purpose.
- This only applies for (F, G, H) fixed. The **total number of such triples** is

$$\lesssim e^{CN}.$$

Therefore the union over all (F, G, H) of all bad events has tiny probability.

The Return Times Theorem

(Bourgain; Demeter-Lacey-Tao-Thiele)
The return times theorem concerns almost-everywhere existence of limits

$$\lim_{N\to\infty} N^{-1} \sum_{k=1}^{N} f(T^{k(\omega)}(x))g(S^{k(\omega)}(y))$$

where T, S are unrelated measure-preserving transformations on two different spaces X, Y.

- The set of good values of x has full measure, and is **universal**; it depends on f but works for every dynamical system (Y, S) and every g.
- The first result of this type was due to Bourgain and applied only to f, g in certain combinations of L^p spaces.
- Demeter-Lacey-Tao-Thiele proved the extension to all $f \in L^p$ and $g \in L^q$ with $p \in (1, \infty]$ and $q \geq 2$.

The case $M = 0$ has an ergodic-theoretic consequence, for return times of sparse random subsequences. Let $(X, \mathcal{A}, T, \mu)$ be any nonatomic dynamical system with probability measure μ.

Theorem

Let $0 \le \gamma < \frac{1}{2}$. Almost every random sequence $\{n_k(\omega)\}$ constructed as above has the this property: Let $p \in (1, \infty]$ and $q \ge 2$. For each $f \in L^p(X)$ there exists a subset $X_0 \subset X$ of full measure such that for every dynamical system $(Y, \mathcal{F}, \nu, \sigma)$, every $g \in L^q(Y)$, and every $x \in X_0$,

$$\lim_{N\to\infty} N^{-1} \sum_{k=1}^{N} f(T^{n_k(\omega)}(x))g(S^{n_k(\omega)}(y))$$

exists for ν-almost every $y \in Y$.